BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

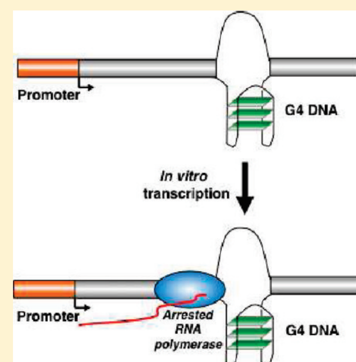
pubs.acs.org/biochemistry

Transcription Arrest by a G Quadruplex Forming-Trinucleotide Repeat Sequence from the Human c-myb Gene

Christopher Broxson, Joshua Beckett, and Silvia Tornaletti*

Department of Anatomy and Cell Biology, University of Florida, College of Medicine and UF Genetics Institute, 1600 SW Archer Road, Gainesville, Florida 32610, United States

ABSTRACT: Non canonical DNA structures correspond to genomic regions particularly susceptible to genetic instability. The transcription process facilitates formation of these structures and plays a major role in generating the instability associated with these genomic sites. However, little is known about how non canonical structures are processed when encountered by an elongating RNA polymerase. Here we have studied the behavior of T7 RNA polymerase (T7RNAP) when encountering a G quadruplex forming-$(GGA)_4$ repeat located in the human c-myb proto-oncogene. To make direct correlations between formation of the structure and effects on transcription, we have taken advantage of the ability of the T7 polymerase to transcribe single-stranded substrates and of G4 DNA to form in single-stranded G-rich sequences in the presence of potassium ions. Under physiological KCl concentrations, we found that T7 RNAP transcription was arrested at two sites that mapped to the c-myb $(GGA)_4$ repeat sequence. The extent of arrest did not change with time, indicating that the c-myb repeat represented an absolute block and not a transient pause to T7 RNAP. Consistent with G4 DNA formation, arrest was not observed in the absence of KCl or in the presence of LiCl. Furthermore, mutations in the c-myb $(GGA)_4$ repeat, expected to prevent transition to G4, also eliminated the transcription block. We show T7 RNAP arrest at the c-myb repeat in double-stranded DNA under conditions mimicking the cellular concentration of biomolecules and potassium ions, suggesting that the G4 structure formed in the c-myb repeat may represent a transcription roadblock in vivo. Our results support a mechanism of transcription-coupled DNA repair initiated by arrest of transcription at G4 structures.



Several naturally occurring DNA sequences can assume alternative (non-B) DNA structures as a result of transient DNA structural changes occurring during normal cellular metabolism.[1] Non canonical DNA structures have been implicated in numerous cellular transactions.[2−6] In addition, non-B DNA structures have also been recently shown to correspond to genomic regions particularly susceptible to genetic instability.[7−13] Transcription appears to play a central role in formation of these structures and in the mutagenesis associated with these genomic sites.[12,14,15] It was shown that the Z-DNA forming sequence $(CG)_{14}$ induces large deletions in mammalian cells in a replication-independent manner.[16] Quadruplex DNA formation was reported to be associated with transcribed G-rich regions located in the immunoglobulin heavy chain S regions undergoing somatic hypermutation.[17] Furthermore, mutagenic G-rich regions with characteristic formation of extended RNA/DNA hybrids have been described in *Saccharomyces cerevisiae* mutants depleted of the THO/TREX complex, which is involved in mRNP processing, and in chicken DT40 cells and human HeLa cells depleted of the ASF/SF2 RNA splicing factor.[18−21] Although these findings suggest a correlation between transcription and non-B DNA induced mutagenesis, little is known about the effect of non-B DNA structures on the transcription process. Transcription may facilitate recruitment of repair proteins at sites of unusual structures as a result of transcription arrest and initiation of transcription coupled repair, a subpathway of nucleotide excision repair initiated at lesions that block progression of RNA polymerase II in vivo.[22]

To understand how non-B DNA structures are processed by the transcription machinery, we have started to characterize the behavior of RNA polymerases when encountering these unusual structures. We have previously shown that a mutagenic quadruplex DNA-forming sequence from the murine Sμ IgG repeat is a strong block to T7 and RNA polymerase II transcription when located in the nontranscribed strand.[23,24] The transcription block is dependent on formation of a G loop, a structure generated by transcription through the G-rich region that consists of a RNA/DNA hybrid in the transcribed strand and a G-rich single-stranded region in the nontranscribed strand.[17]

To establish direct correlations between the presence of G4 DNA in the transcription template and its inhibitory effects on transcription, we have utilized an experimental approach that takes advantage of the ability of T7 RNA polymerase to transcribe single-stranded templates and of the absolute requirement of KCl for formation and stability of G quartets, core structural

Received: February 10, 2011
Revised: April 5, 2011
Published: April 06, 2011

ACS Publications © 2011 American Chemical Society

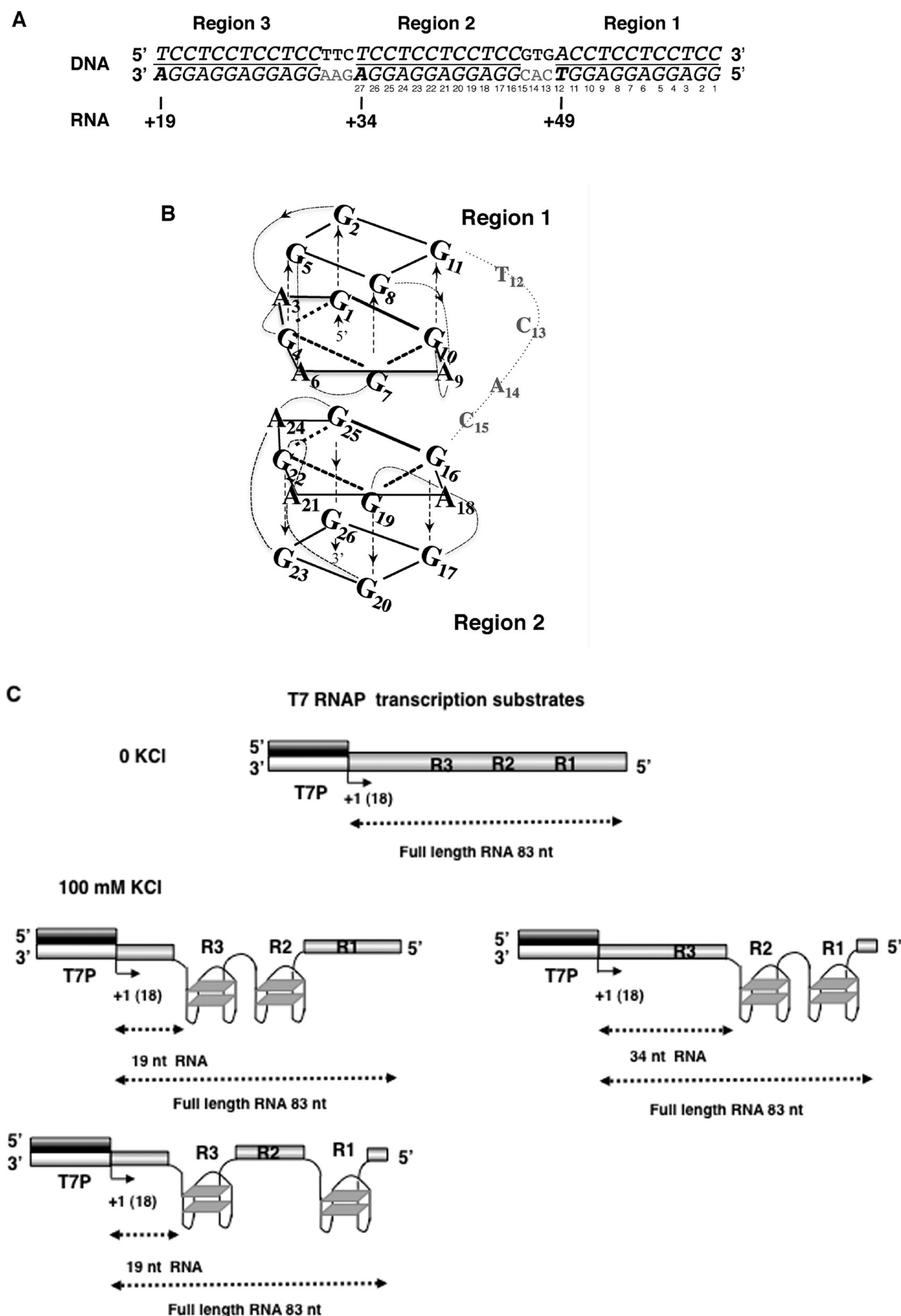


Figure 1. DNA sequence used in this study. (A) $(GGA)_4$ repeat region from the c-myb proto-oncogene; the three $(GGA)_4$ repeats that fold into G4 structure are underlined: Region 1 (R1), Region 2 (Region 2), and Region 3 (R3); the location and size of the RNAs resulting from transcription arrest at the c-myb repeat are marked on the sequence; the DNA bases spanning Region 1 and Region 2 are numbered 1 to 27. (B) Schematic representation of a tetrad/heptad/heptad/tetrad (T:H:H:T) G quadruplex DNA structure formed between any two regions of the c-myb GGA repeat sequence. Bases are numbered according to the sequence of Region 1 and Region 2 described in panel A. (C) Transcription substrates and expected RNA products. DNA substrates for T7 transcription are obtained by annealing a 17 mer containing the minimal T7 promoter sequence with a 100 mer containing a sequence complementary to the T7 promoter followed by the c-myb G4-forming sequence at position +19 from the transcription start site. Annealing of the two oligonucleotides in the presence or absence of KCl generates transcription substrates in which the G4-forming sequence has folded or not into G4 quadruplex structure. Inhibitory effects on transcription caused by formation of the alternative DNA structure are indicated by synthesis of short RNAs compared to full-length products. Three T:H:H:T G quadruplex DNA structures can form in the c-myb repeat: R3-R2, R3-R1, and R2-R1. RNA sizes expected from transcription arrest at these structures are indicated by arrows.

components of G4 DNA.[25] Utilizing this approach, we have studied the behavior of T7 RNA polymerase when encountering a well characterized G4-forming sequence located in the promoter of the human c-myb proto-oncogene[26,27] (Figure 1).

Table 1. Sequences Used in This Study (5′-3′)

T7 promoter

5′-TAATACGACTCACTATA-3′

c-myb100ts

5′-CGTAGAGGATCTCAGGAGAAAGAGGAGGAGGAGGTCACGGAGGAGGAGGAGAAGGAGGAGGAGGAAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

c-myb100R1ts

5′-CGTAGAGGATCTCAGGAGAAAGAGCACGAGCAGCTCACGGAGGAGGAGGAGAAGGAGGAGGAGGAAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

c-myb100R2ts

5′-CGTAGAGGATCTCAGGAGAAAGAGGAGGAGGAGGTCACGCAGCAGCAGCAGAAGGAGGAGGAGGAAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

c-myb100R3ts

5′-CGTAGAGGATCTCAGGAGAAAGAGGAGGAGGAGGTCACGGAGGAGGAGGAGAAGCAGCAGGAGCAAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

c-myb100R1-R2ts

5′-CGTAGAGGATCTCAGGAGAAAGAGCAGCAGCACGTCACGCAGCACGAGCAGAAGGAGGAGGAGGAAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

c-myb100R1-R3ts

5′-CGTAGAGGATCTCAGGAGAAAGAGCAGCAGCAGCTCACGGAGGAGGAGGAGAAGCAGCAGCAGCAAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

c-myb100R2-R3ts

5′-CGTAGAGGATCTCAGGAGAAAGAGGAGGAGGAGGTCACGCAGCAGCAGCAGAAGCAGCAGCAGCAAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

c-myb100ts

5′-CGTAGAGGATCTCAGGAGAAAGAGGAGGAGGAGGTCACGGAGGAGGAGGAGAAGGAGGAGGAGGAAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

c-myb100Uts

5′-TAATACGACTCACTATAGGGAGATATCAACCTGTTTCCTCCTCCTCCTTCTCCTCCTCCTCCGTGACCTCCTCCTCCTCTTTCTCCTGAGA-TCCTCTACG-3′

GTG100ts

5′-CGTAGAGGATCTCTCGAGTCTTCTTCTGTGCACTCTTCTTCTGGATCCACAGGACGGGTGTGGTAAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

GTG100Uts

5′-TAATACGACTCACTATAGGGAGATATCAACCTGTTTACCACACCCGTCCTGTGGATCCAGAAGAAGAGTGCACAGAAGAAGACTCGAGAGA-TCCTCTACG-3′

c-myb100nts

5′-*TAATACGACTCACTATA*GGGAGATATCAACCTGTTAGGAGGAGGAGGAAGAGGAGGAGGAGGCACTGGAGGAGGAGG -
AGAAAGAGGAGAGATCCTCTACG-3′

c-myb100Lnts

5′-CGTAGAGGATCTCTCCTCTTTCTCCTCCTCCTCCAGTGCCTCCTCCTCCTCTTCCTCCTCCTCCTAACAGGTTGATATCTCCC*TATAGTGA-GTCGTATTA*-3′

EXPERIMENTAL PROCEDURES

Proteins and Reagents. T7 RNAP was purchased from Promega. Proteinase K was from Invitrogen. DNA oligonucleotides were purchased from Midland. Highly purified NTPs were purchased from Amersham Pharmacia Biotech. Radiolabeled nucleotides were from MP Biomedical. Piperidine and dymethyl sulfate (DMS) were from Sigma. Restriction enzymes were from New England Biolabs.

Preparation of DNA Templates for Transcription. Synthetic DNA templates for transcription reactions with T7 RNAP consisted of single-stranded or double-stranded DNA oligonucleotides. Single-stranded DNA substrates consisted of DNA oligonucleotides in which a double-stranded T7 promoter region was generated by annealing a 23 mer oligonucleotide of sequence complementary to the T7 promoter sequence at the 3′ end of the single-stranded oligonucleotide. The oligomers sequences are listed in Table 1. T7 transcription substrates were obtained by incubating 50 pmol of T7 primer with 50 pmol of 100 mer oligonucleotide in a final volume of 10 μL for 3 min at 95 °C, followed by slow cooling down to 37 °C, and further incubation overnight at 37 °C. After the annealing step, samples were kept at −20 °C. To induce G4 DNA formation, which is dependent on the presence of potassium ions,[25] the annealing reaction was carried out in 100 mM KCl. Control samples were incubated in 100 mM NaCl or 100 mM LiCl or in TE buffer. Similarly, double-stranded DNA substrates were obtained by annealing 50 pmol of c-myb100ts oligonucleotides with the complementary c-myb100Uts

in the presence or absence of 100 mM KCl and concentrations of polyethyleneglycol 200 up to 40%.[28]

T7 RNAP Transcription Reactions. For single round transcription reactions, the DNA templates (1 picomol) were incubated at 37 °C for 5 min in a mixture of 50 units of T7 RNAP, 40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 2 mM spermidine, 10 μCi [α-^{32}P] GTP, 10 mM dithiothreitol, 212 units of RNAsin, 200 μM ATP and UTP. Elongation proceeds in the presence of ATP, UTP, and [α-^{32}P] GTP until T7 RNAP reaches the end of the C-less cassette (nucleotide 7), at which the first CTP is necessary for incorporation. Heparin was added at a final concentration of 250 μg/mL to prevent further initiation[29−31] and 200 μM CTP, UTP, and GTP was added to allow elongation to continue. Incubation continued at 37 °C for 30 min. Reaction conditions for multiple round transcription differed from those for single round only in that all four nucleotides were included from the beginning of the transcription reaction at a final concentration of 200 μM ATP, CTP, UTP, 20 μM GTP, and 10 μCi [α-^{32}P] GTP. Reactions were stopped by addition of 5 μg of proteinase K, 1% SDS, 100 mM TrisHCl (pH 7.5), 50 mM EDTA, and 150 mM NaCl, followed by incubation for 15 min at room temperature. The nucleic acids were precipitated with ethanol, resuspended in formamide dye, and denatured at 90 °C for 3 min. The transcription products were resolved on a 5% denaturing polyacrylamide gel in Tris borate-EDTA containing 7 M urea. Gels were dried and autoradiographed using intensifying screens.

Transcripts were quantified using a Typhoon phosphorimager and ImageQuant software from GE Healthcare. The extent of arrest at R3, R2, or R1 was calculated by dividing the intensity of the 19 nt, 34 nt, or 49 nt transcript band by the sum of the intensity of the arrested and runoff RNA bands. For multiple round transcription experiments, the extent of arrest was adjusted by the G content of the RNA transcripts.

Dimethyl Sulfate Protection Assays. 2.5 pmol of 100 mers containing the G4 forming sequence were end-labeled with [γ-^{32}P] ATP and polynucleotide kinase. They were then heat denatured at 95 °C followed by slow cooling down to room temperature in TE buffer in the presence or absence of 100 mM KCl. The 100 mers were cooled down to 0 °C before being added to a solution containing 50 mM sodium cacodylate (pH 7.0) 1 mM EDTA pH 8.0. Five μL of dymethyl sulfate were added and left to react for 5 min at room temperature. The reactions were stopped by addition of 1.5 M sodium acetate, pH 7.0, 1 M β-mercaptoethanol, 1 mg/mL tRNA. DNA was ethanol precipitated and resuspended in 1 M piperidine. After cleavage at 90 °C for 30 min, reactions were stopped by chilling in ice followed by ethanol precipitation. The samples were resuspended in 100 μL of water and dried overnight in a speed vac concentrator. Samples were resuspended in 4 μL of formamide dye followed by denaturation for 3 min at 90 °C. The DNA samples were separated on a 12% denaturing polyacrylamide gel in Tris borate-EDTA containing 8.3 M urea.

RESULTS

G4-Forming Sequences from the c-myb Proto-oncogene Block T7 Transcription Elongation in a K^+-Dependent Manner. We have studied the effect of G4-forming sequences on transcription elongation utilizing an approach that takes advantage of the ability of T7 RNA polymerase to synthesize RNA transcripts from single-stranded DNA molecules, when a double-stranded promoter region is provided to initiate transcription, and of the property of G4 DNA to form in single-stranded DNA in the presence of KCl, which stabilizes G quartets.[25,32] By using this approach, we could promote the transition to G4 DNA independently of the topology of the DNA substrate and of the presence of active transcription along the template DNA. Furthermore, by utilizing substrates in which the G4-forming region was single-stranded we could eliminate the possibility that any inhibitory effect on transcription resulted from formation of alternative DNA structures other than G4 DNA, which is unique in its ability to form in single-stranded G-rich regions in the presence of KCl. Therefore, by using this approach we could make direct correlations between arrest of transcription and presence of this structure in the DNA substrate. Any inhibitory effect on T7 transcription would be indicated by synthesis of RNAs shorter than the full-length runoff transcripts (Figure 1C). Furthermore, since we know the location of the G4-forming-sequence in our templates and we can map the size of the transcription products, we can carefully determine where transcription was arrested with respect to the non-B DNA structure.

Using this approach, we have studied the behavior of T7 RNA polymerase when transcribing a well characterized G4-forming sequence located at position +17 downstream of the transcription start site of the human c-myb proto-oncogene[26,27] (Figure 1). It was recently shown that this G4-forming sequence also plays a regulatory role in c-myb gene expression.[27] Therefore, it was relevant to study the effect of this G4-forming-sequence on transcription. This c-myb sequence consists of three $d(GGA)_4$ repeats: Region 3 (R3), Region 2 (R2), and Region 1 (R1) (Figure 1A). The structure of a $d(GGA)_4$ repeat has been previously determined by nuclear magnetic resonance under physiological K^+ conditions,[26,27] where it was shown that $d(GGA)_4$ folds into a intramolecular G quadruplex consisting of a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G(:A) heptad (Figure 1B). Two intramolecular quadruplexes form a dimer which is stabilized through stacking interactions between the heptads of the two quadruplexes (Figure 1B) to form a tetrad/heptad/heptad/tetrad (T:H:H:T) quadruplex structure. On the basis of these structural studies, it has been proposed that any two of the three $d(GGA)_4$ repeats located in the c-myb gene can fold into a T:H:H:T structure (Figure 1B).[26,27] Evidence in support of this model has been recently obtained by circular dicroism spectroscopy, DNA polymerase, and RNA polymerase arrest assays and dimethyl sulfate protection assays.[27]

Transcription templates were generated by annealing a 17 nt long DNA oligomer corresponding to the minimal T7 promoter sequence, to a 100 nt long DNA oligomer including the T7 promoter sequence at the 3′ end followed by the c-myb GGA repeat located 19 nt downstream of the T7 RNA polymerase transcription start site (Figure 1 and Table 1). To promote the transition to G4 DNA structure, the 17 nt and 100 nt DNA oligomers were annealed in the presence of 100 mM KCl. Control samples were incubated under the same conditions except that KCl was omitted from the annealing reaction. DNA samples consisting of an identical DNA sequence except that the G4-forming sequence was substituted with a random DNA sequence devoid of adjacent G residues to impede G4 DNA formation were annealed under the same conditions.

As a negative control for our transcription studies, we utilized GTG 100ts (Table 1), a sequence that we had previously shown does not affect transcription elongation when located in a plasmid construct downstream of the T7 promoter.[30] When

Figure 2. T7 RNA polymerase transcription on DNA substrates containing the $(GGA)_4$ repeat region from the c-myb gene. DNA templates were transcribed in vitro such that the transcripts were radioactively labeled. Elongation was allowed to proceed for 30 min at 37 °C after addition of NTPs to the reaction. Lanes 1 and 2: Control transcription template (GTGts) obtained after annealing of a 17 mer and a 100 mer in the absence (lane 1) or presence (lane 2) of KCl as described in Materials and Methods. Lanes 6 and 7: Transcription template obtained after annealing of a 17 mer and a 100 mer in the absence (lane 6) or presence (lane 7) of 100 mM KCl to induce formation of G4 structure in the c-myb repeat as detailed in Materials and Methods. M: 10 bp DNA ladder. Lanes 3−5: 18, 19, and 20 nt long RNAs synthesized from transcription of a 35 mer, 36 mer, and 37 mer oligonucleotide DNA template of identical sequence as that of the 100 mer c-myb containing substrate were ran on a 12% denaturing polyacrylamide gel in parallel with RNAs generated from transcription of the 100 mer c-myb repeat containing substrate. M: 10 bp DNA ladder. Sites of transcription arrest are marked with arrows. RO: runoff RNA.

the T7 transcription substrate containing this sequence was annealed either in the absence or presence of 100 mM KCl, T7 RNA polymerase preceded until the end of the GTG 100ts DNA template, resulting in formation of full-length RNA transcripts (Figure 2, lanes 1−2). In addition, several shorter RNAs were also observed, likely resulting from a decrease in T7 RNA polymerase processivity when transcribing a single-stranded substrate. However, no difference in the pattern of RNA transcripts was observed when the transcription substrates were annealed in the presence or absence of KCl. When the T7 transcription substrate containing the c-myb repeat was annealed in the absence of KCl, transcription proceeded until the end of the DNA template, resulting in formation of full-length RNAs (Figure 2, lane 6). However, when the T7 transcription substrate containing the c-myb repeat was annealed in the presence of 100 mM KCl, the amount of full length RNA decreased dramatically, with a concomitant increase of two shorter RNAs (Figure 2, lane 7). These RNAs corresponded to 19 nt and 34 nt long transcripts, as determined by running them in parallel with RNA markers 18, 19, and 20 nt long, obtained from transcription of DNA substrates 35, 36, and 37 nt long of sequence identical to that of the c-myb DNA substrate (Figure 2, lanes 3−5). These RNA sizes are expected if T7 transcription is arrested 1 nt into Region 3 and 1 nt into Region 2, respectively (Figure 1C).

Dimethyl Sulfate Protection Assays Implicate G4 DNA Formation in the Mechanism of Transcription Arrest at the c-myb $(GGA)_4$ Repeat. Our finding that T7 arrest at the c-myb repeat was dependent on the presence of KCl and that it occurred on the single-stranded substrates suggested that the c-myb $(GGA)_4$ repeat had folded into G4, as previously observed by others.[27] To corroborate this possibility, we carried out dimethyl sulfate (DMS) footprinting experiments, which are routinely used to identify G4 DNA formed in single-stranded synthetic oligonucleotides in vitro.[25] This assay is based on the property of DMS to attack the N7 of guanine, which is accessible to DMS methylation when it is present in single-stranded or duplex DNA but not when it is paired to the exocyclic amino group of a neighboring guanine in a G-quartet.[33] As a result, the Gs in the quartet are protected from methylation and subsequent piperidine cleavage compared to the unstructured control. Protection from DMS methylation is visualized as a decrease in intensity or absence of the corresponding G band after DNA separation on denaturing polyacrylamide gels. When the oligonucleotide containing the c-myb $(GGA)_4$ repeat was treated with DMS and piperidine, followed by 12% denaturing polyacrylamide gel electrophoresis (Figure 3, lane 1), we observed bands of similar intensity corresponding to cleavage at the Gs in the three GGA repeats located in R1, R2, and R3. However, when the c-myb $(GGA)_4$ repeat was incubated with 100 mM KCl before DMS treatment, we observed a significant decrease in the intensity of the Gs corresponding to the GGA repeat located in R1 and R2 (Figure 3, lane 2), indicating that these guanines where partially protected from DMS modification. We did not observe DMS protection of the Gs located in the $(GGA)_4$ repeat of Region 3, a result also observed by others,[27] suggesting that under these conditions, the G4 structure in R3 may be formed but is unstable. This was confirmed by performing DMS footprinting on oligonucleotides in which G to C substitutions where introduced in R1 or R2 to prevent G4 formation. Under these conditions, we observed a significant decrease in the intensity of the Gs corresponding to the GGA repeat located in R1 and R3 (Figure 3, lane 4), or in R2 and R3 (Figure 3, lane 6), indicating that these guanines where partially protected from DMS modification.

Mutational Analysis of the c-myb $(GGA)_4$ Repeats Reveals a Correlation between Polymerase Arrest and Quadruplex Structure Formation. To further characterize the contribution of structure formation to RNA polymerase arrest, we have introduced mutations in the $(GGA)_4$ repeats that prevent G4 DNA formation, followed by transcription assays. If G4 DNA formation was responsible for transcription arrest at the c-myb repeat, we expected to see that changing every G into C in anyone of the three $(GGA)_4$ repeats in the c-myb sequence, would result in changes in the pattern and extent of transcription arrest depending on which repeat was mutated. Indeed, we found that when every G was mutated into C in region R1 or R2, and, as a result, only regions R3-R2 or R3-R1 could fold into G4 structure (Figure 1B), only the RNA corresponding to arrest at A19, the first base of the $(GGA)_4$ repeat in R3, was observed (Figure 4A, lanes 7−8). When every G was mutated into C in region R3, and, as a result, only region R2-R1 could fold into G4, only the RNA corresponding to arrest at A34, the first base of the $(GGA)_4$ repeat in R2 was observed. The extent of transcription arrest at R3 (19 nt RNA) increased from ~70% to ~90% in the

Figure 3. Dimethyl sulfate protection assay. DMS methylation patterns are indicated by gray circles (partially methylated) and dark circles (methylated). The wild type GGA repeat (lane 2) shows partial protection from DMS methylation in R1 and R2 in the presence of potassium. The R2 mutant (lane 4) shows partial protection from DMS methylation in R1 and R3 in the presence of potassium. The R1 mutant (lane 6) shows partial protection from DMS methylation in R2 and R3 in the presence of potassium.

R1 mutant, and from ~70% to ~95% in the R2 mutant (Figure 4B). The extent of arrest at R2 (34 nt RNA) increased from ~25% to ~90% in the R3 mutant. Control transcription reactions carried out on the same DNA substrates preincubated in the absence of KCl (Figure 4A, lanes 1−5) or in the presence of 100 mM LiCl (Figure 4A, lanes 11−15), ionic conditions that prevent G4 DNA formation in this repeat, produced mostly full-length RNA, as expected. These results further suggest that G4 DNA formation is implicated in the mechanism of arrest at the c-myb $(GGA)_4$ repeat.

A Single $(GGA)_4$ Repeat Is Sufficient to Cause Transcription Arrest. To assess whether the presence of one out of three $(GGA)_4$ repeats was sufficient to cause transcription arrest, transcription substrates were prepared in which every G was mutated into C in two out of three $(GGA)_4$ repeats. If a T:H module was sufficient for G4 DNA formation, we expected to see arrest even when only one out of three repeats was present.[26] Transcription was carried out on these substrates and compared to transcription on the wt sequence (Figure 5). We found that an intact region R3, region R2 or Region 1 was sufficient to cause transcription arrest, as indicated by formation of a 19 nt, a 34 nt, or a 49 nt RNA respectively (Figure 5A, lanes 2−4). The extent of transcription arrest at R3 (19 nt RNA) increased from ~70% to ~90% in the R1-R2 mutant. The extent of arrest at R2 (34 nt RNA) increased from ~25% to ~80% in the R1-R3 mutant. The 49 nt RNA, generated from transcription of the R2-R3 mutant, corresponded to ~85% arrest (Figure 5B).

The Two K^+ Dependent Transcription Blocks Observed in the c-myb $(GGA)_4$ Repeat Are Sites of Transcription Arrest and Not Transient Transcription Pauses. To determine whether the two sites of transcription stalling observed after preincubation of the DNA substrates with KCl represented a transient pause or were instead an absolute block to RNA polymerase progression, we have followed RNA synthesis as a function of time under single round or multiple round conditions.[24] If the GGA repeat represented a permanent block to T7 progression, we expected to see no change between RNA species as a function of time, indicating that once the RNA polymerase reached the structure, it could not continue transcribing. If the GGA repeat was a transient pause instead, we expected to see that the amount of nascent transcripts decreased as a function of time, with corresponding increase of longer RNA species corresponding to read-through products. We found that the relative amount of nascent transcript R2 and R3, corresponding to arrest at the GGA repeat in region 2 or in region 3, as well as the runoff RNA, did not significantly change over time up to 30 min under single or multiple round conditions (Figure 6). This result indicated that the two sites of transcription stalling represented an absolute block to transcription elongation. We calculated that the total amount of blockage at the c-myb $(GGA)_4$ triplet repeat sequence reached 70% under either single or multiple round transcription conditions.

Transcription Arrest in the c-myb $(GGA)_4$ Repeat Occurs in Double-Stranded Substrates under Conditions Mimicking the Concentration of Biomolecules and Potassium Ions Found in Cells. We have shown that under physiological concentrations of potassium ions the c-myb $(GGA)_4$ repeat represents a strong block to T7 elongation. To closely reproduce the in vivo situation, we have extended this analysis to test whether the potassium dependent-inhibitory effect on transcription by the c-myb $(GGA)_4$ repeat would also be observed in double-stranded DNA templates. When DNA is double-stranded, the B DNA form is favored over the quadruplex DNA conformation. In agreement with these predictions, we found that when we annealed the $(CCT)_4$ repeat containing strand to the $(GGA)_4$ repeat strand in the presence of 100 mM KCl (Figure 7, lane 2) transcription proceeded through the $(GGA)_4$ repeat, suggesting that under these conditions G4 DNA did not form in the c-myb

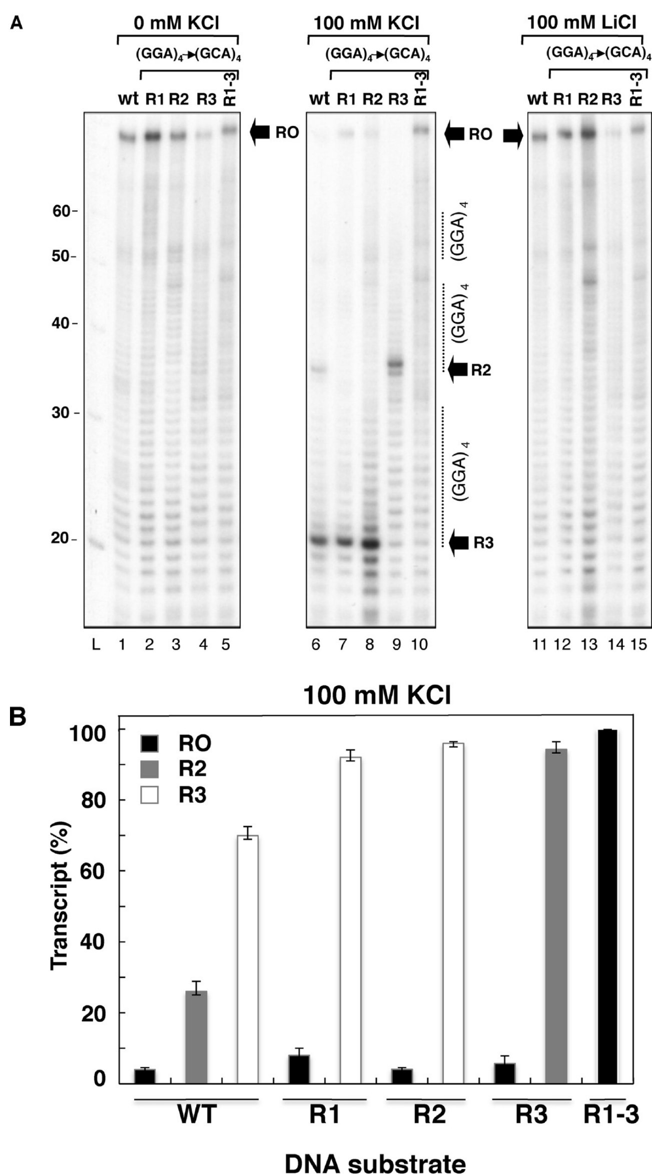


Figure 4. T7 RNA polymerase transcription on DNA substartes containing GGA → GCA mutations in R1, R2, or R3: (A) T7 RNA polymerase transcription was carried on DNA substrates containing the wt c-myb repeat (wt) or the c-myb repeat containing GGA → GCA mutations in region 1 (R1), region 2 (R2), region 3 (R3), or R1, R2, R3 (R1−3). DNA substrates were annealed in the absence of KCl (lanes 1−5), in the presence of 100 mM KCl (lanes 6−10), or in the presence of 100 mM LiCl (lanes 11−15). Sites of transcription arrest are marked with arrows. (B) Quantitation of the transcription results from panel A, lanes 6−10.

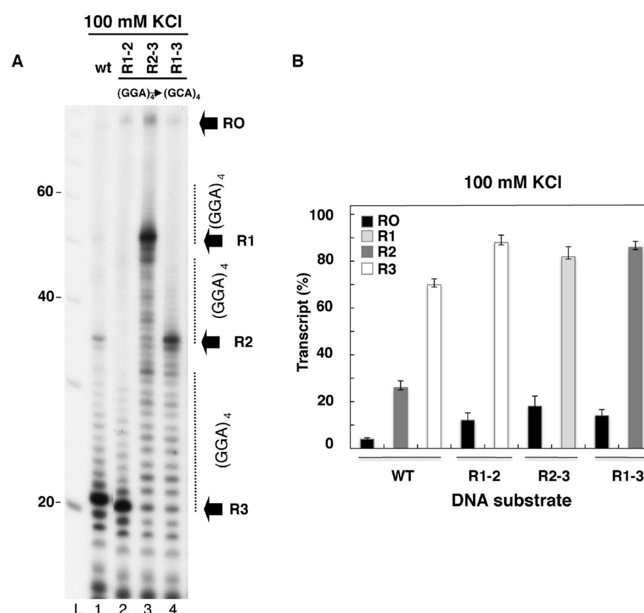


Figure 5. T7 RNA polymerase transcription on DNA substrates containing GGA → GCA mutations in R1-2, R1-3 or R2-3. (A) T7 RNA polymerase transcription was carried on DNA substrates containing the wt c-myb repeat (wt) or the c-myb repeat containing GGA → GCA mutations in region 1 and 2 (R1-2), region 1 and 3 (R1-3) or region 2−3 (R2-3). L: 10 bp DNA ladder. Sites of transcription arrest are marked with arrows. (B) Quantitation of the transcription results from panel A.

repeat. To promote the structural transition to G4 DNA under conditions mimicking the intracellular environment, which is characterized by high concentration of macromolecules (10−40% of the total cellular volume), we have added increasing concentrations of PEG 200 during the annealing of the DNA complementary strands, followed by transcription. PEG 200 is a molecular crowding agent that favors transition from duplex to quadruplex DNA by excluded volume effects.[28,34] We found that addition of PEG 200 alone during annealing did not have any effect on RNA polymerase progression, suggesting that G4 DNA did not form in the c-myb repeat (Figure 7, lanes 3, 5, 7, 9). However, when increasing concentrations of PEG 200 up to 40% were added in the presence of 100 mM KCl during the annealing step, we observed formation of two short transcripts, corresponding to arrest at R3 and R2, in addition to full-length RNA (Figure 7A), as previously shown when transcription was carried out with the single-stranded T7 substrate (Figure 2). Quantitation of these transcripts indicated that the extent of arrest at R2 and R3 in the double-stranded substrates reached 70% (Figure 7B). Under the same experimental conditions, the presence of the c-myb $(GGA)_4$ repeat in the nontranscribed strand did not affect transcription arrest by T7 RNA polymerase either in the presence or absence of KCl and PEG 200, as indicated by synthesis of only full-length RNAs (data not shown).

To confirm that the transcription templates were double-stranded downstream of the G4 repeat and not an hybrid of ds DNA at the promoter region followed by G4 DNA in one strand and the other one unpaired, we have prepared dsDNA substrates preincubated or not with KCl and PEG 200, to induce the G4 DNA structure, and we have digested them with restriction enzyme BfucI, which cuts DNA downstream of the c-myb $(GGA)_4$ repeat, before running transcription (Figure 8). We reasoned that if the DNA was double-stranded downstream of the G4 structure the restriction enzyme BfucI would cut the DNA, resulting in synthesis of a 76 nt runoff RNA, 7 nt shorter than the runoff RNA obtained from transcription of the undigested substrate (Figure 8A). On the other hand, if the DNA was not double-stranded downstream of the G4 structure, BfucI would not cut, thus resulting in synthesis of the same size 83 nt runoff RNA as the BfucI-uncut DNA substrate. We found that transcription of the BfucI cut DNA generated mostly the 76 nt runoff RNA, and only a small fraction of the 83 nt runoff RNA,

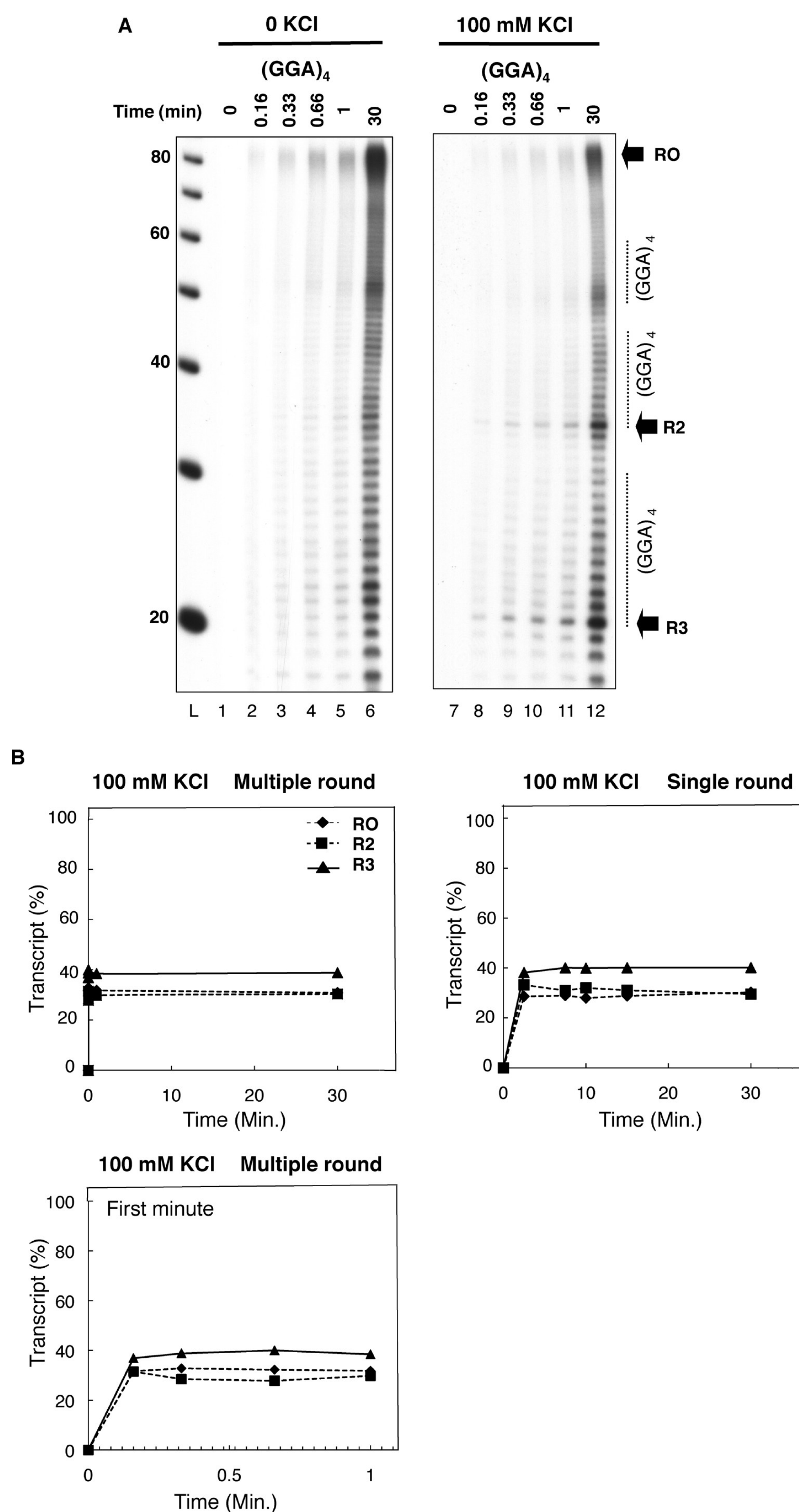


Figure 6. Time course of T7 RNA polymerase transcription on DNA substrates containing the GGA repeat region from the c-myb gene. (A) DNA substrates were annealed in the absence (lanes 1−6) or in the presence of KCl (lanes 7−12); the gel shows RNAs generated from T7 transcription under multiple round conditions; L: 10 bp ladder; RO: runoff transcript; R3, RNA arrested at region 3; R2: RNA arrested at region 2. (B) Quantitation of the transcription results obtained under single round or multiple round conditions (see text for details).

which represented less than 10% of the total runoff RNA (76 nt +83 nt RNA, Figure 8, lanes 4 and 12), confirming that the transcription substrates where mostly double-stranded downstream of the G4 structure.

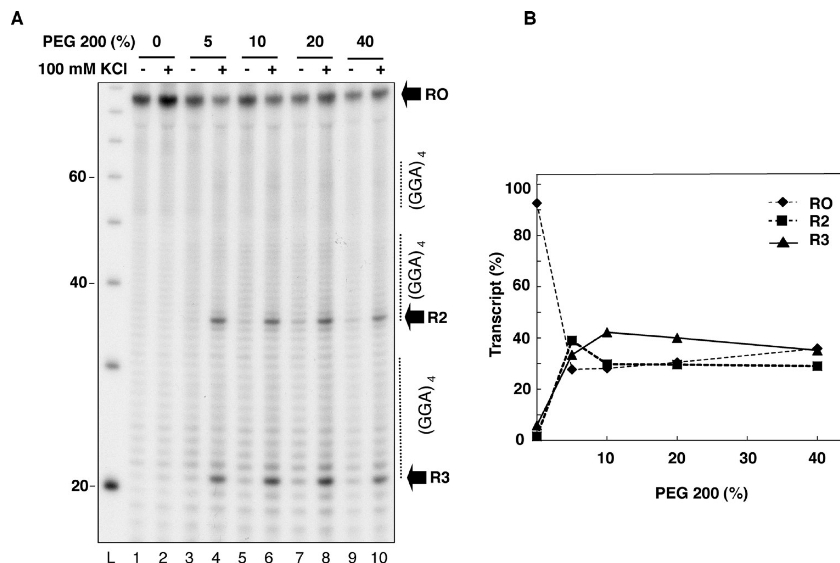


Figure 7. T7 RNA polymerase transcription on double-stranded DNA substrates containing the $(GGA)_4$ repeat region from the c-myb gene. (A) The c-myb repeat-containing DNA oligonucleotide and the complementary strand were heated at 95 °C followed by slow cooling down to room temperature in the presence (lanes 2, 4, 6, 8, 10) or absence (lanes 1, 3, 5, 7, 9) of 100 mM KCl, without (lane 1) or with (lanes 4, 6, 8, 10) increasing concentrations of PEG 200. L: 10 bp DNA ladder. RO: runoff transcript; R3, RNA arrested at region 3; R2: RNA arrested at region 2. (B) Quantitation of the transcription results.

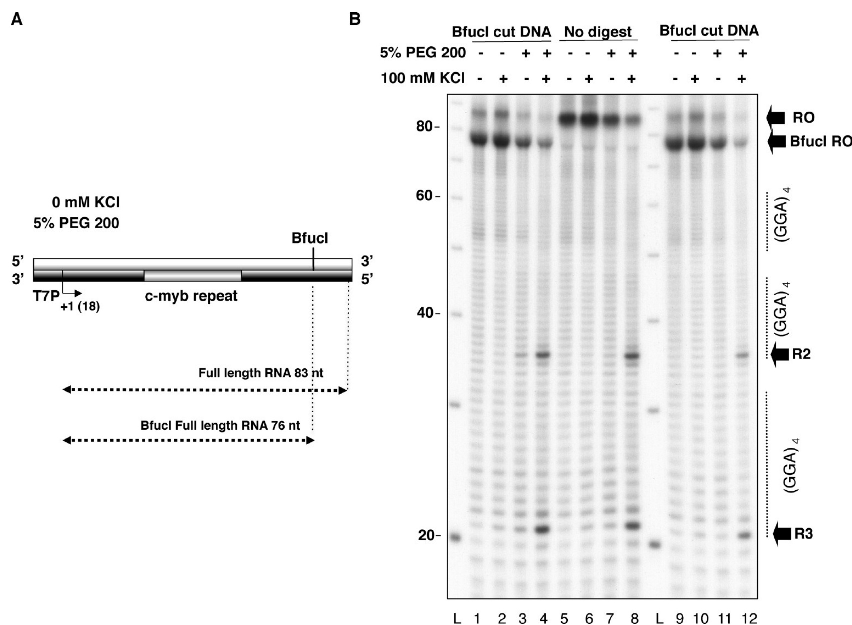


Figure 8. T7 RNA polymerase transcription on double-stranded DNA substrates digested with restriction enzyme BfucI. (A) Double-stranded DNA substrates obtained by annealing the complementary strands in the presence or absence of 100 mM KCl and 5% PEG 200 were digested (lanes 1−4 and 9−12) or not (lanes 5−8) with restriction enzyme BfucI prior to T7 RNA polymerase transcription. BfucI restriction digest was carried out in the reaction buffer recommended by the manufacturer (lanes 1−4) (NEB buffer 4: 20 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM dithiothreitol (pH 7.9)) or in a reaction buffer not containing potassium, to prevent any transitions to G4 DNA during the incubation, in which BfucI digests DNA with 100% efficiency (lanes 9−12) (NEB buffer 1: 10 mM Bis Tris Propane-HCl, 10 mM MgCl2, 1 mM dithiotheitol (pH 7.0)). (B) Graphic representation of the double stranded DNA template. BfucI: Location of the BfucI restriction site. The full length RNAs generated from transcription of the BfucI cut (76 nt BfucI RO) or uncut DNA substrate (83 nt RO) are marked with double-ended arrows.

■ DISCUSSION

We have studied the effect of a G4-forming $(GGA)_4$ repeat sequence from the c-myb proto-oncogene on transcription elongation by T7 RNA polymerase. In order to establish direct correlations between the presence of this non-B DNA structure and its effect on transcription, we have taken advantage of the unique ability of T7 RNA polymerase to transcribe single-stranded DNA templates and of the requirement of G4 DNA

formation on single-strandedness and on the presence of potassium ions.[35] Assembly of the T7 transcription initiation complex occurs on a transcription template composed of a double-stranded T7 promoter and of a single-stranded DNA region, containing a G quadruplex-forming sequence downstream of the transcription start site. To promote G4 DNA formation and to stabilize G quartets, core structural components of G4 DNA, T7 transcription substrates were assembled in the presence of KCl.[25]

Using this approach, we show that a G4-forming $(GGA)_4$ repeat from the human c-myb proto-oncogene is a strong block to transcription by T7 RNAP. Two K^+ dependent T7 arrest sites were observed that mapped at the beginning of the first (Region 3) and second (Region 2) $(GGA)_4$ repeat (Figure 2). Consistent with formation of G4 DNA in the c-myb $(GGA)_4$ repeat, we found partial protection from DMS methylation of the G residues located in Region 1 and Region 2 in the wt sequence (Figure 3). DMS protection was observed in Region 3 when base substitutions where introduced to prevent G4 DNA formation in Region 2 or Region 1 of the transcription templates, suggesting that under these conditions the G4-forming sequence in Region 3 is the less stable, as previously reported by others.[26,27,36] Further evidence that G4 DNA formed in our substrates was the finding that substituting KCl with LiCl during the annealing reaction did not have any significant inhibitory effect on T7 transcription (Figure 4A). In addition, when we introduced base substitutions in the GGA repeat sequences that are expected to prevent G4 DNA formation we found that they eliminated the transcription block (Figures 4 and 5). Our results are consistent with formation of a tetrad/heptad/heptad/tetrad (T:H:H:T) G quadruplex structure in the wild type c-myb sequence between Region 3/2, Region 3/1, and Region 2/1 of the c-myb GGA repeat, as indicated by the presence of the 19 and 34 nt RNA and not of the 49 nt RNA (Figure 2). However, we cannot exclude the possibility that the lack of obvious stalling at R1 in the WT sequence may be simply because the majority of transcripts were blocked at R3 and R2 before reaching R1. It is unlikely that structures other than G4 DNA may have been responsible for the observed inhibitory effect on transcription since T7 arrest was observed when transcription was proceeding on single-stranded DNA molecules in the presence of KCl. In addition, the dependence of T7 arrest on the presence of potassium ions strongly suggests that a triplex or a hairpin structure was not the cause of arrest in the double-stranded template, since formation of these alternative DNA structures is independent of the presence of K^+.[36]

The location of the arrest sites at the beginning of the GGA repeats suggests that the G4 structure represents a physical barrier to RNA polymerase progression. This conclusion is further supported by our findings that the extent of transcription arrest did not change with time under single or multiple round transcription conditions (Figure 6).

When double-stranded substrates were annealed under conditions mimicking the physiological concentrations of potassium ions and biomolecules, both sites of transcription arrest where observed, that mapped at the same location as those detected in single-stranded DNA, suggesting that G4 DNA formation in the transcription substrates was responsible for the inhibitory effect on transcription. Our results represent the first direct evidence showing that a G4 forming-sequence is a block to transcription in double-stranded DNA. This effect was observed when double-stranded substrates were annealed in the presence of 100 mM KCl and concentrations of PEG 200 up to 40%. These PEG concentrations are very similar to the physiological concentration of biomolecules (30−40%, w/v) in living cells,[28,37] suggesting that the c-myb $(GGA)_4$ repeat has the potential to fold into G4 in vivo. The observation that this sequence is located in the transcribed strand 17 nt downstream of the c-myb gene transcription start site raises the possibility that formation of G4 DNA in the c-myb G repeat may directly control c-myb gene expression by causing arrest of transcription elongation in vivo. However, it is difficult to extrapolate the data obtained with the viral polymerase to the eukaryotic enzyme. For this reason, we will extend our studies to characterize the effect of the c-myb repeat on transcription of the mammalian RNA polymerase.

When double-stranded substrates were utilized as transcription templates, we found that transcription arrest occurred only when the c-myb repeat was located in the transcribed strand (Figures 7 and 8 and data not shown). Interestingly, the presence of the c-myb repeat in the nontranscribed strand did not have any detectable effect on RNA polymerase progression (data not shown). This result suggests that the c-myb repeat may act as a roadblock to polymerase movement along the DNA, impeding addition of the next nucleotide when it reaches the catalytic site, thus promoting polymerase arrest. This effect of the c-myb repeat on transcription is reminiscent of the effect of DNA lesions on transcription elongation and may implicate possible similarities in the mechanism of transcription arrest at non B DNA compared to that for DNA lesions. Only when lesions are located in the transcribed strand of template DNA they affect RNA polymerase progression, suggesting that the presence of the lesion at or near the catalytic site, where nucleotide addition occurs, plays a critical role in transcription arrest. Similarly, the presence of a G quartet at or near the T7 catalytic site may represent an insurmountable barrier for the polymerase. On the other hand, the presence of the structure in the nontranscribed strand may not be as critical for the progression of the transcription complex, although they may have a significant effect on the DNA structure. Similarly, cisplatin-induced cross-links and benzopyrene diol-epoxide-induced lesions only affect transcription when they are located in the transcribed strand although they cause significant effects on DNA structure.[38,39]

Our findings that the c-myb G repeat is a strong block to transcription when it is located in the transcribed strand raises the possibility that this sequence may be recognized by repair proteins as another form of endogenous DNA damage thus initiating a futile cycle of repair in an otherwise undamaged DNA region. This reiterative and futile repair cycle may generate mutations when occurring in a highly transcribed gene, due to the natural error rate of repair enzymes. This mechanism of gratuitous TCR has been proposed as one of the mechanisms involved in generating genomic instability associated with genomic regions with potential to assume unusual DNA structures in vivo.[12,22,24]

AUTHOR INFORMATION

Corresponding Author

*Tel.: 352-273-8935; fax 352-846-1248; e-mail: silviat@ufl.edu.

Funding Sources

This work was supported by the Florida Department of Health Bankhead-Coley Cancer Research Program (NIR 08BN07) and by funds provided by the Department of Anatomy and Cell Biology, UF College of Medicine and by the UF Genetics Institute.

ABBREVIATIONS USED

G4 DNA, G quadruplex DNA; T7 RNAP, T7 RNA polymerase; DMS, dimethyl sulfate

REFERENCES

(1) Sinden, R. R. (1994) *DNA Structure and Function*, Academic Press, San Diego.
(2) Phan, A. T., Kuryavyi, V., and Patel, D. J. (2006) DNA architecture: from G to Z. *Curr. Opin. Struct. Biol. 16*, 288–298.
(3) Maizels, N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol. 13*, 1055–1059.
(4) Fry, M. (2007) Tetraplex DNA and its interacting proteins. *Front. Biosci. 12*, 4336–4351.
(5) Wang, G., and Vasquez, K. M. (2007) Z-DNA, an active element in the genome. *Front. Biosci. 12*, 4424–4438.
(6) Kouzine, F., and Levens, D. (2007) Supercoil-driven DNA structures regulate genetic transactions. *Front. Biosci. 12*, 4409–4423.
(7) Bacolla, A., Wojciechowska, M., Kosmider, B., Larson, J. E., and Wells, R. D. (2006) The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA Repair 5*, 1161–1170.
(8) Wang, G., and Vasquez, K. M. (2006) Non-B DNA structure-induced genetic instability. *Mut. Res. 598*, 103–119.
(9) Raghavan, S. C., and Lieber, M. R. (2006) DNA structures at chromosomal translocation sites. *BioEssays 28*, 480–494.
(10) Maizels, N. (2008) Genomic stability: FANCJ-dependent G4 DNA repair. *Curr. Biol. 18*, 613–614.
(11) Mirkin, S. M. (2007) Expandable DNA repeats and human disease. *Nature 447*, 932–940.
(12) Lin, Y., Hubert, J., L., and Wilson, J. H. (2009) Transcription destabilizes triplet repeats. *Mol. Carc. 48*, 350–361.
(13) Wang, G., and Vasquez, K. M. (2009) Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. *Mol. Carc. 48*, 286–298.
(14) Liu, L., and Wang, J. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U.S.A. 84*, 7024–7027.
(15) Aguilera, A., and Gâomez-Gonzâalez, B. (2008) Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Gen. 9*, 204–217.
(16) Wang, G., Christensen, L. A., and Vasquez, K. M. (2006) Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A. 103*, 2677–2682.
(17) Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., and Maizels, N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev. 18*, 1618–1629.
(18) Huertas, P., and Aguilera, A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell 12*, 711–721.
(19) Li, X., and Manley, J. L. (2005) Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell 122*, 365–378.
(20) Li, X., and Manley, J. L. (2005) New talents for an old acquaintance: the SR protein splicing factor ASF/SF2 functions in the maintenance of genome stability. *Cell Cycle 2*, 1706–1708.
(21) Lin, Y., Dent, S. Y., Wilson, J. H., Wells, R. D., Napierala, M. R loops stimulate genetic instability of CTG.CAG repeats. *Proc. Natl. Acad. Sci. U.S.A. 107*, 692—697.
(22) Hanawalt, P. C., and Spivak, G. (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell. Biol. 9*, 958–970.
(23) Tornaletti, S., Park-Snyder, S., and Hanawalt, P. C. (2008) G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J. Biol. Chem. 283*, 12756–12762.
(24) Tornaletti, S. (2009) Transcriptional processing of G4 DNA. *Mol. Carc. 48*, 326–335.
(25) Huppert, J. L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev. 37*, 1375–1384.
(26) Martsugami, A., Ouhashi, K., Kanagawa, M., Liu, H., Kanagawa, S., Uesugi, S., and Katahira, M. (2001) An intramolecular quadruplex of $(GGA)_4$ triplet repeat DNA with G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol. 313*, 255–269.
(27) Palumbo, S. L., Memmott, R. G., Uribe, D. J., Krotova-Khan, Y., Hiurley, L. H., and Ebbinghaus, S. W. (2008) A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic Acids Res. 36*, 1755–1769.
(28) Zheng, K. W., Chen, Z., Hao, Y. H., and Tan, Z. (2009) Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. *Nucleic Acids Res. 38*, 327–338.
(29) Viswanathan, A., and Doetsch, P. W. (1998) Effects of non-bulky DNA base damages on *Escherichia coli* RNA polymerase-mediated elongation and promoter clearance. *J. Biol. Chem. 273*, 21276–21281.
(30) Tornaletti, S., Patrick, S. M., Turchi, J. J., and Hanawalt, P. C. (2003) Behavior of T7 RNA polymerase and mammalian RNA polymerase II at site-specific cisplatin adducts in the template DNA. *J. Biol. Chem. 278*, 35791–35797.
(31) Ferrari, R., Rivetti, C., and Dieci, G. (2004) Transcription reinitiation properties of bacteriophage T7 RNA polymerase. *Biochem. Biophys. Res. Commun. 315*, 376–380.
(32) He, B., Kukarin, A., Temiakov, D., Chin-Bow, S. T., Lyakhov, D. L., Rong, M., Durbin, R. K., and McAllister, W. T. (1998) Characterization of an unusual, sequence-specific termination signal for T7 RNA polymerase. *J. Biol. Chem. 273*, 18802–18811.
(33) Sen, D., and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature 334*, 364–366.
(34) Zhou, J., Wei, C., Jia, G., Wang, X., Tang, Q., Feng, Z., and Li, C. (2008) The structural transition and compaction of human telomeric G-quadruplex induced by excluded volme effect under cation-deficient conditions. *Biophys. Chem. 136*, 124–127.
(35) Temiakov, D., Anikin, M., and McAllister, W. T. (2002) Characterization of T7 RNA polymerase transcription complexes assembled on nucleic acid scaffolds. *J. Biol. Chem. 277*, 47035–47043.
(36) Usdin, K. (1998) NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases. *Nucleic Acids Res. 26*, 4078–4085.
(37) Zimmerman, S. B. (1993) Macromolecular crowding effects on macromolecular interactions: some implications for genome structure and function. *Biochim. Biophys. Acta 1216*, 175–185.
(38) Scicchitano, D. (2005) Transcription past DNA adducts derived from polycyclic aromatic hydrocarbons. *Mutat. Res. 577*, 146–154.
(39) Tornaletti, S. (2009) Transcription-coupled DNA repair: directing your effort where it's most needed. *Cell. Mol. Life Sci. 66*, 1010–1020.